Title: Open Data for Nuclear Explosion Monitoring (NEM)

Author(s): Stead, Richard J.

Intended for: EIVR58 virtual meeting Oct 11-15

Issued: 2021-10-08

Los Alamos NATIONAL LABORATORY

Open Data for Nuclear Explosion Monitoring (NEM)

Richard Stead, EES-17

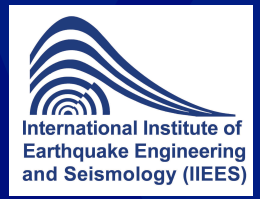
2021/10/13

Outline

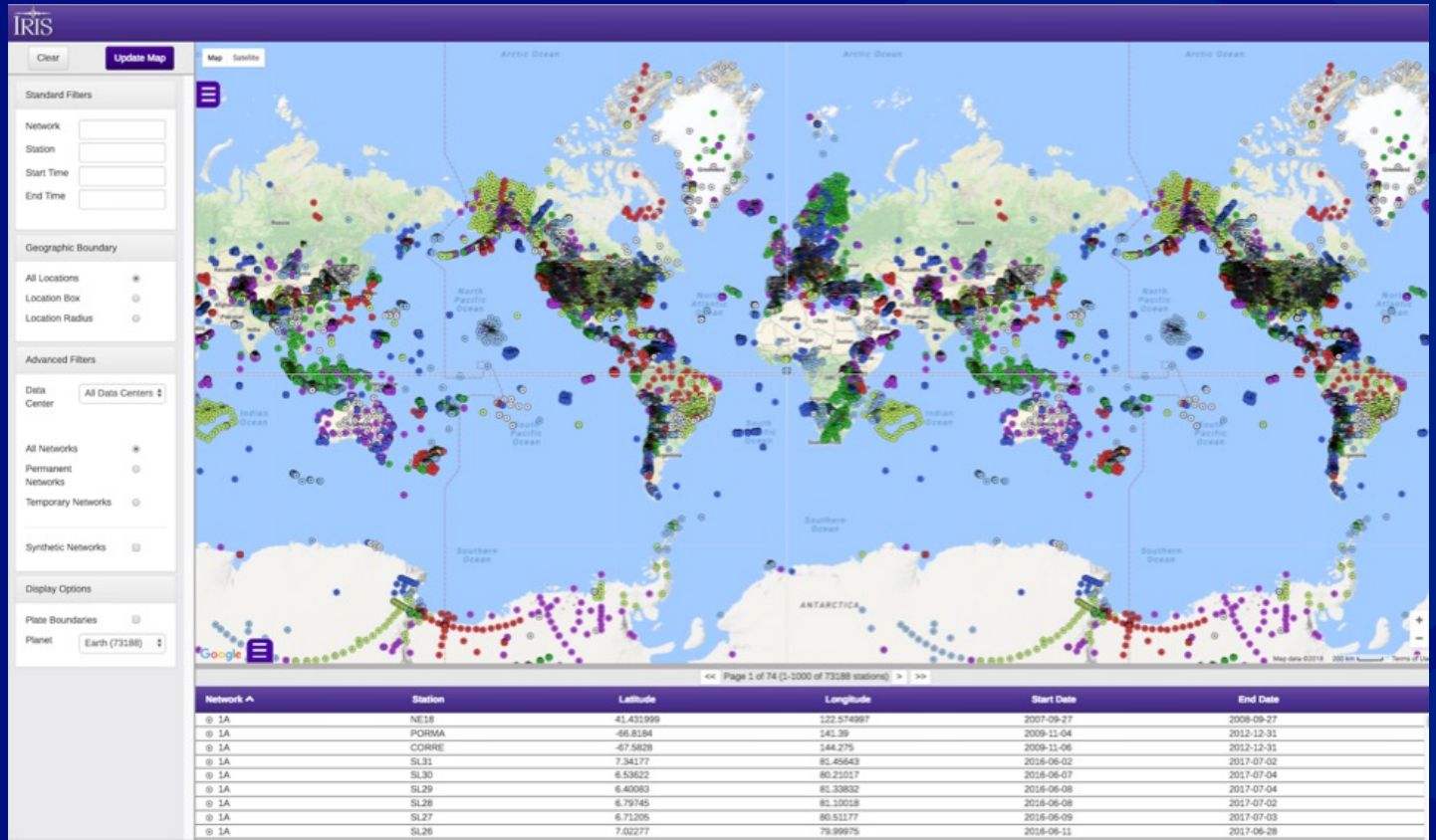
- Seismic waveform data
  - Well-known sources
  - Less well-known sources
  - Access
  - NA-22 sponsored
  - Legacy
- Seismic bulletin data (brief discussion)
- Summary

Seismic waveform data: sources

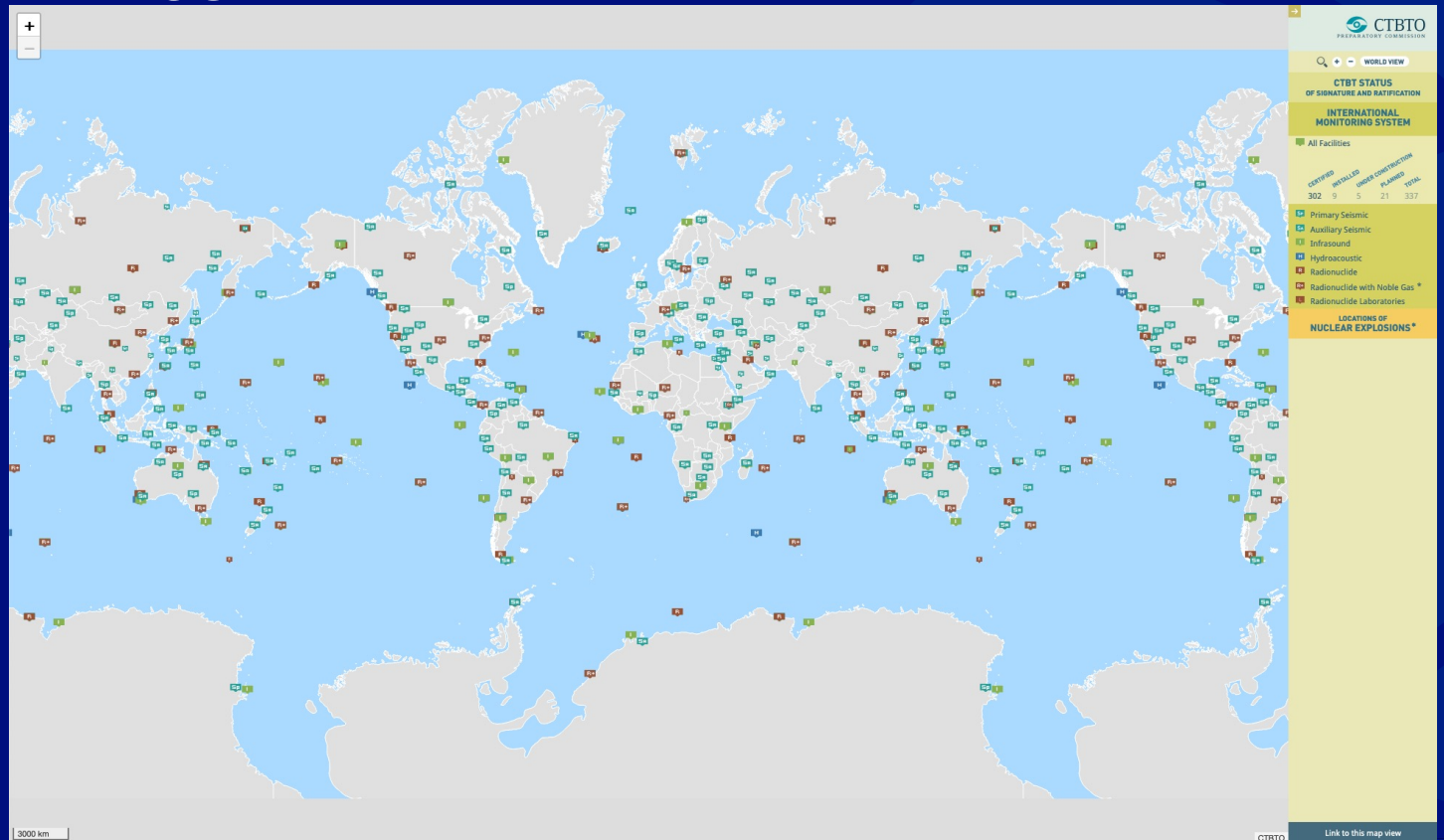
- Well-known national and international
  - US: IRIS
  - Europe: Orfeus, Geofon, …
  - CTBTO IDC
- Other countries with general access
  - Japan: JMA, NIED Hi-NET, F-NET
  - South Korea: KMA, KIGAM
  - Iran: IIEES, University of Tehran
  - …



Seismic waveforms: data available through IRIS and other FDSN-affiliated data centers



Seismic waveforms: CTBTO PrepCom data available through NDCs



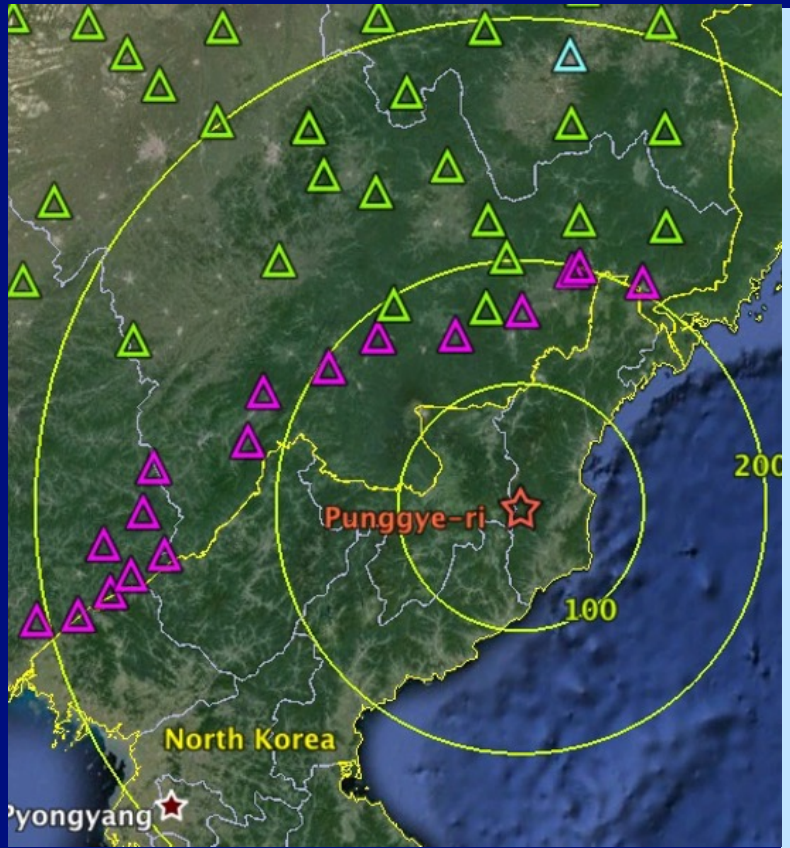
Seismic waveform data: Issues with well-known sources

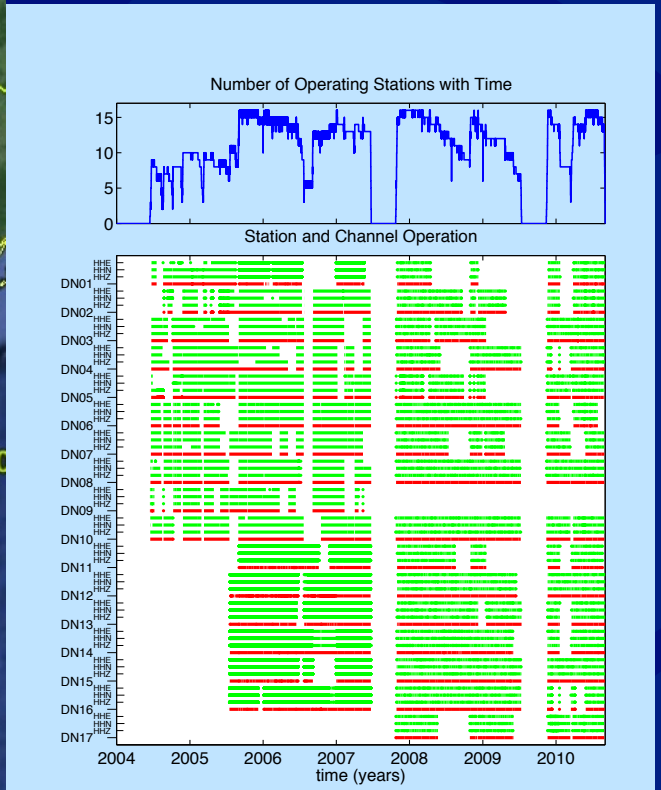
- The data sources tend to have the highest quality data and metadata, particularly for more recent data sets
- Early data from sources such as IRIS tend to have some metadata issues.
  - Standards such as SEED may be interpreted differently by different data providers
  - Station naming and channel naming can be confusing, particularly local practice versus FDSN standards
  - Station location information can be problematic, especially when comparing to the station registry at the ISC
  - In some rare instances, problems with the waveforms themselves have been found, such as undocumented gain changes or problems with instrumentation
  - IRIS has tools such as Mustang to help identify issues with data
- There can be such a thing as too much data, at least for some applications

Seismic waveform data: other sources

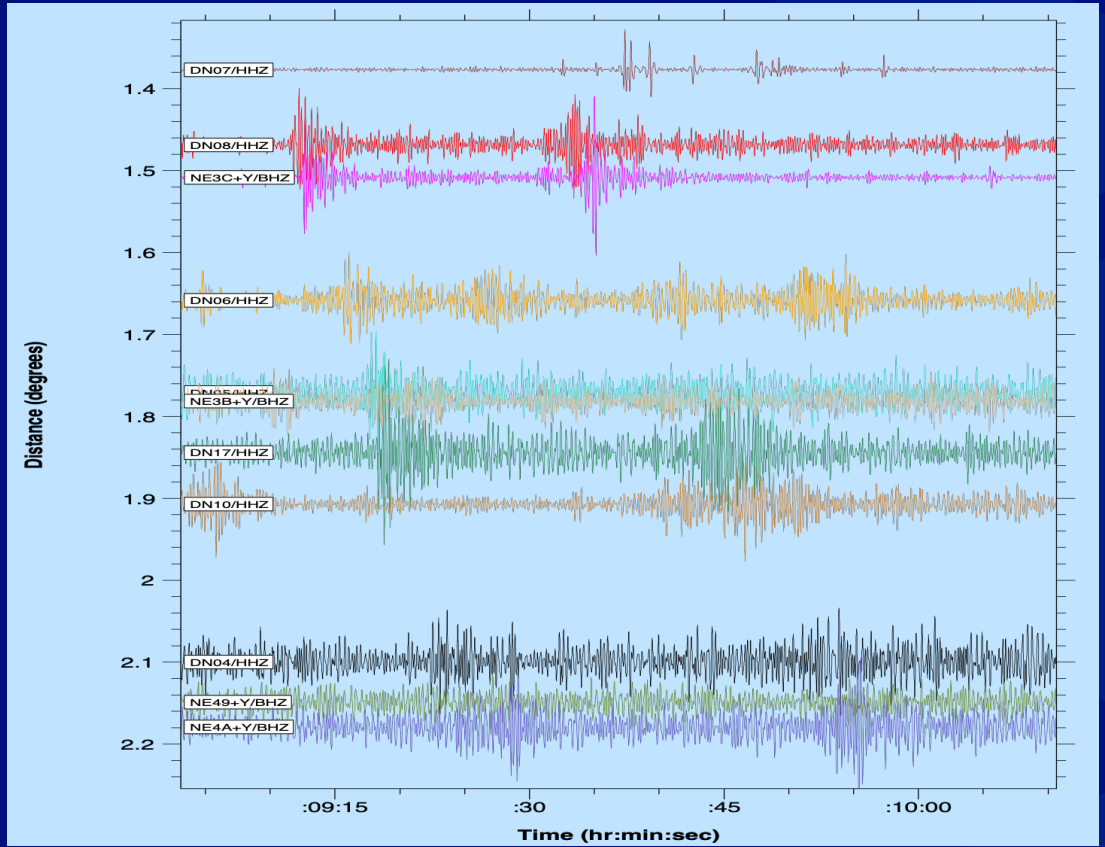
- University researchers
  - Data sets independently released
    - Example: the Dongbei data released by LDEO
  - Data sets available through collaborations or funded joint research
    - Researchers may be sensitive about release -> treat as OUO
- Corporations
  - Data sets available through funded joint research
    - Corporations may be sensitive about release (proprietary data) -> treat as OUO
- Government
  - Data collected under US government funding are legally the property of the US government and are therefore available to any US government agency with the necessary clearance and need-to-know.
    - Note that this does not always happen in practice due to lack of enforcement
- Regional Networks
  - US examples: SCEDC, University of Utah, UC Berkeley, University of Nevada Reno

Seismic Waveforms: Dongbei





Seismic waveforms: Dongbei example (May 2010)



Seismic waveform data: Issues with less well-known sources

- These sources of data tend to have a great deal more data and metadata issues than well-known sources
- Metadata may be almost completely absent or contain gross errors
- The waveforms have been observed to contain numerous problems such as spikes/glitches/dropouts/clipping, sudden changes in response, and timing errors
- There tend to be many gaps in the data

Seismic waveform data: access

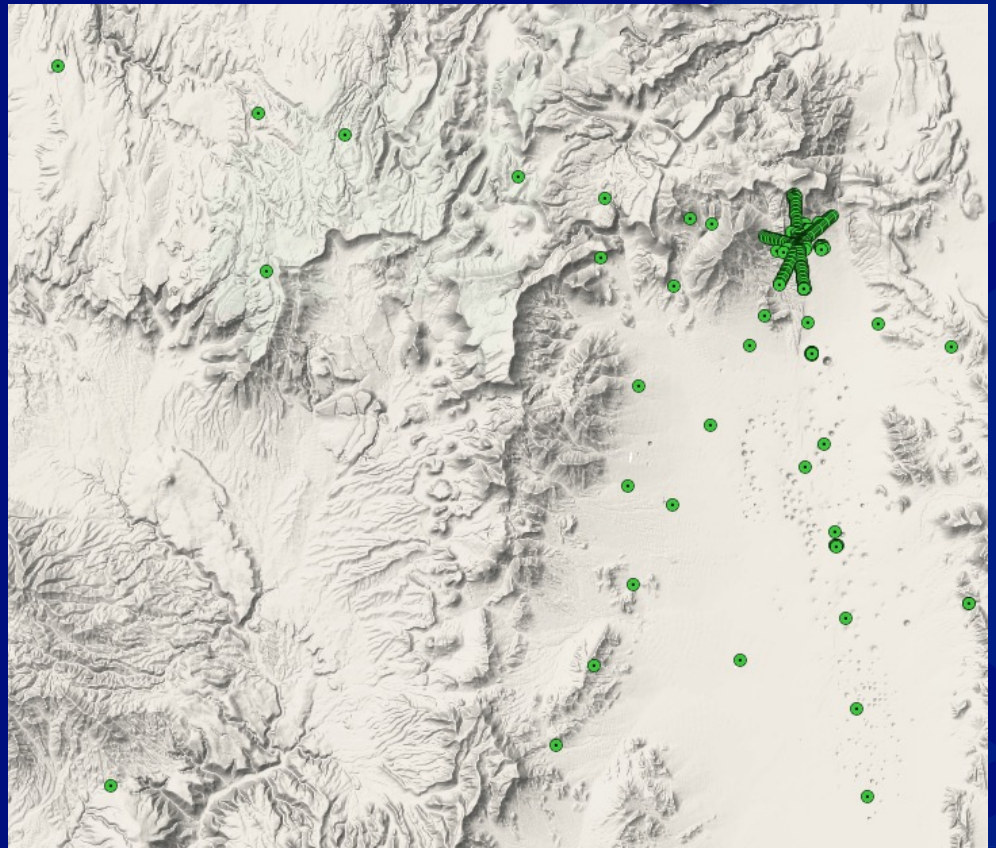
- Common methods (mostly for well-known sources and country sources)

  Note: some sites may require some sort of registration process
  - Web-based queries
    - Example: IRIS' SeismiQuery
  - Email-based queries
    - Examples: AutoDRM, breq_fast
  - FTP sites
- Other (common for collaborations, government data, etc)
  - Shipment of disk drives
  - Dropbox and similar sites
  - Cloud storage

Seismic waveform data: NA-22 sponsored

- SPE-DAG
  - Very large data set intended to better understand the generation of seismic waves at the source of an explosion
  - Public release of data through IRIS
- LYNM (PE1, et al.)
  - Experiment design takes advantage of lessons learned under SPE-DAG
  - At least some data will likely be OUO
  - Public data will be released through IRIS after a holding period
- Vulcan/Helios/PELE/FULL TOSS (and related)
  - Focus is more on weaponization than nuclear testing, includes high-explosive tests
  - Highly multi-modal data sets with somewhat limited and inconsistent seismic data
  - All data considered OUO

Seismic waveforms: SPE



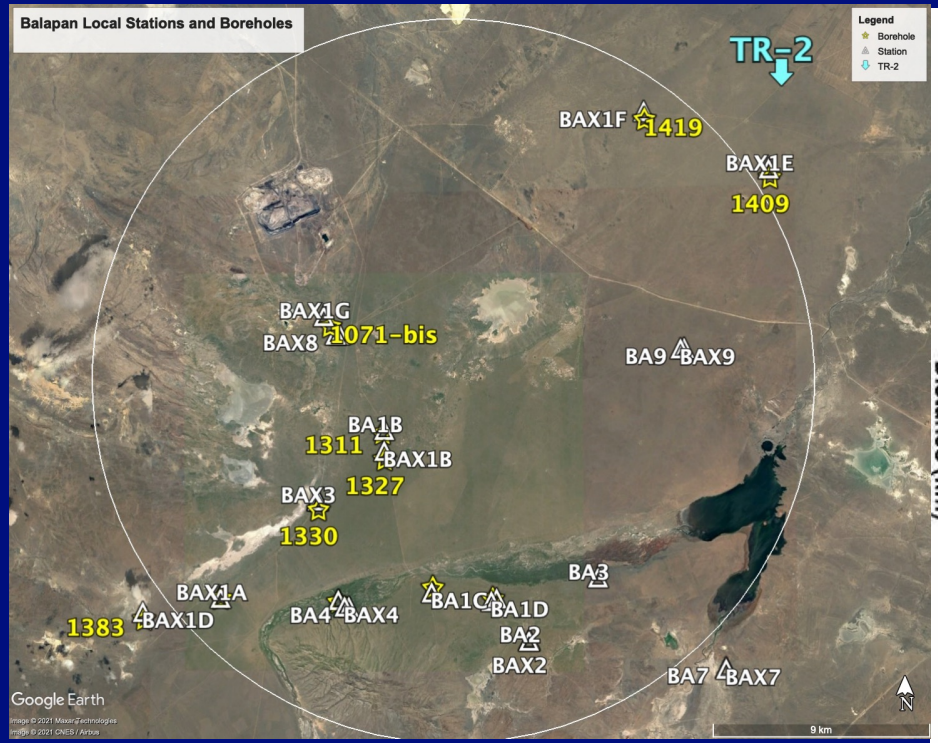
Seismic waveform data: issues with NA-22 sponsored data

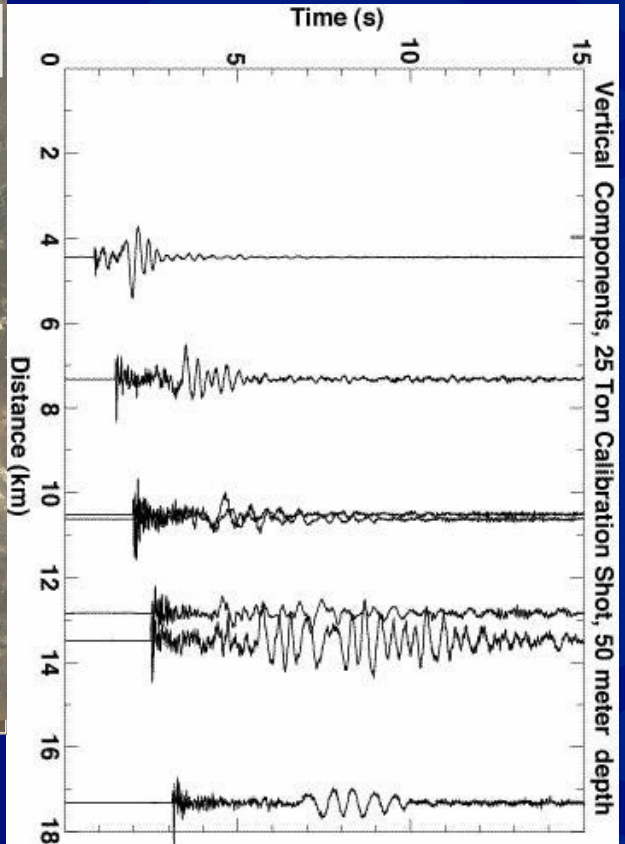
- NA-22 sponsored data are relatively new, and benefit from higher standards in the community in general
- NA-22 large venture data products do show a learning curve
  - Early data from SPE may suffer from some minor metadata issues, particularly with regard to instrument response
- Some data types may be more difficult to access
  - Large-n (sometimes called “nodal”) data from SPE can be hard to obtain
- Some datasets may take substantial effort to transfer due to their size

Seismic waveform data: legacy

- US Nuclear test data
  - Many US nuclear tests included seismic sensors (both seismometers and accelerometers) deployed in the near-field
  - Some analog data were digitized long ago, but provenance is sketchy
  - Many analog tapes still exist, but extremely limited ability to digitize
  - Some data from later tests exist in digital formats, but recovery of these is also difficult
- Others
  - Test site data
    - Borehole closure data from Semipalatinsk (1998-1999) recovered from legacy tapes in 2020
  - Soviet DSS and seismic expeditions
    - Efforts by LDEO and University of Saskatchewan have recovered a lot of this in digital form, but metadata issues continue to be a problem

Seismic waveforms: Legacy data from Semipalatinsk





Seismic waveform data: issues with legacy data

- Problems recovering the data in digital form
- Problems determining timing
- Problems reconstructing metadata
  - Station and channel codes
  - Station locations
  - Instrumentation

NOTE: these problems in some cases may be insurmountable, but the data are irreplaceable.

Seismic waveform data: other considerations

Seismic bulletin data: sources

- Well-known:
  - US: USGS
  - Europe: ISC
  - CTBTO IDC
- Others:
  - Kazakhstan: KNDC
  - South Korea: KMA, KIGAM
  - Japan: JMA
  - …

Seismic bulletin data: NNSA sponsored

- ”Siberian” bulletin
  - Michigan State University, with some collaboration from LANL
  - Compiled legacy bulletin data from multiple regional networks
  - Most of this work was completed between 2005 and 2015
- Report NV-209 (official released data on US nuclear tests)
  - See also the Russian “Bear Book” (The Nuclear Tests of the U.S.S.R, V.N. Mikhailov editor)

Summary

- There are a wide variety of sources of open data for NEM, but there are issues:
  - Conditions for data access vary widely; in many cases establishing collaborative relationships is important
  - Methods of data access vary widely from easy-to-use web sites and email protocols to manual transfers often involving exchange of physical media
    - May involve recovery of older data off of legacy media
  - Quality of the data, in particular the metadata varies widely; metadata in some cases is almost completely absent
  - Some modern datasets are massive, and can exceed 100 TB, complicating data transfer
- Data sharing between National Laboratories can be improved
  - Data sources sensitive regarding release of their data must be protected (protect data as OUO)